DOCUMENT RESUME

ED 075 046	LI 004 282
AUTHOR	Heilprin, Laurence B.
TITLE	On Access to Knowledge in the Social Sciences and Humanities, From the Viewpoint of Cybernetics and Information Science.
INSTITUTION	City Univ. of New York, Flushing, N.Y. Queens Coll. Library Science Dept.
PUB DATE	Apr 72
NOTE	34p.; (15 References); Conference on Access to Knowledge and Information in the Social Sciences and Humanities: Problems and Implications, Apr. 5-6, 1972, NYC
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Communication (Thought Transfer); *Cybernetics; Humanities; *Information Science; Information Theory; Models; Social Sciences
IDENTIFIERS	*Information Transfer

ABSTRACT

The literature of knowledge is a very large system in the cybernetic sense of intractibility to control. Improving access to it needs some simplifying theory. A step in this direction is a hypothesis constructed from basic concepts. These include cybernetic concepts of variety and requisite variety; a version of the mathematical concept of homomorphic mapping; and information scientific concepts: an invariant 3-segmented information science (IS) path, and short and long duration modes of message propagation. Since all disciplines are symbiotic, defining a distinct IS domain is purely pragmatic. However, the IS concepts do define a domain, which acts as a reference frame convenient for locating the substructures necessary for cognitive access to literature. The most critical processes in access occur in our minds, not in data files. Access to knowledge requires completing an IS path--connecting two minds across a variable physical segment. The special problems of access to the literature of the social sciences and the humanities are chiefly those of small classes with large variety to overcome. Certain variety-suppressing devices should be particularly helpful at this stage. However, there is a large, long term cost for the disciplines and professions concerned. (A related document is LI 004 283.) (Author/SJ)

Library Science Department, Queens College of the City University of New York

Conference on Access to Knowledge and Information in the Social Sciences and Humanities: Problems and Implications, April 5-6, 1972, New York City

A B S T R A C T

On Access to Knowledge in the Social Sciences and Humanities, From the Viewpoint of Cybernetics and Information Science.

Laurence B. Heilprin

Abstract: The literature of knowledge is a "very large system" in the cybernetic sense of intractibility to control. Improving access to it needs some simplifying theory. A step in this direction is a hypothesis constructed from a small number of basic concepts. These include cybernetic concepts of variety and requisite variety; a version of the mathematical concept of homomorphic mapping; and information-scientific concepts: an invariant 3-segmented IS path, and short and long duration (SD and LD) modes of message propagation. Since all disciplines are symbiotic, defining a distinct IS domain is purely pragmatic. However, the IS concepts do define a domain, which acts as reference frame convenient for locating the substructures necessary and sufficient for cognitive access to literature.

Cognition is visualized as two main processes: data gathering or acquisition of sensory variety, and data processing or homomorphic "abstracting". Understanding is suggested as pattern recognition of the spectrum of abstractions into which sensory experience is decomposed. A simple model is given for a treelike (minimum class extension) abstract "level" structure. It is based on discrete quanta of variety (stored direct sense impressions, SSI's) and "natural" associative processes, to which access is had by Pavlovian "conditionally" associated stimuli (symbols). A model of information search of an LD collection incorporates these basic processes: sensing, pattern recognition of prior abstractions, and alternation of two homomorphic mappings-- decompression and selection (decision making on surrogate or image collections). Some familiar IS devices are interpreted as regulators of variety and its flow. The most critical processes in access occur in our minds, not in data files, libraries or computers. Access to knowledge requires completing an IS path-- connecting two minds across a variable physical segment.

The special problems of access to the literature of the social sciences and the humanities are chiefly those of small classes with large variety to overcome. Certain variety-suppressing devices such as thesauri should be particularly helpful at this stage. However, there is a large, long term cost ahead for the disciplines and professions concerned.

On Access to Knowledge in the Social Sciences and Humanities, From the Viewpoint of Cybernetics and Information Science

Laurence B. Heilprin
School of Library and Information Services
and
Computer Science Center, University of Maryland

The subject of this conference: access to knowledge in the social sciences and humanities, forms part of the larger subject, access to knowledge. This in turn is part of a still larger subject, human mind-to-mind communication. The last is an immense field which touches and borrows from every science. Nevertheless a central core of what we are after is found in information science or "informatics" (IS), and in cybernetics (C), a somewhat more abstract discipline which goes beyond human communication to include communication in general, animal or machine. The two sciences may be roughly distinguished by the fact that in cybernetics communication is the <u>means</u> to regulate and control systems of any kind, whereas in information science communication is the <u>end</u> in itself, and is chiefly confined to human mind-to-mind communication.

The reason for including our subject within broader classes is simple. Human thinking is performed in terms of classes, and a system of any kind is simply a class located within a larger class -- its environment. The processes of the system may be[1] divided into what have been called its "internal transformations", and those processes which cross its boundaries and relate it to its environment. Therefore a theory of access to knowledge in the social sciences and humanities includes two kinds of relations: those within the subject, and those relating the subject to its immediate environment, knowledge, and to the larger environment, mind-mind communication. Since our everyday experience tends to make us more familiar with internal than with internal-external relationships, the first and larger part of this paper is about the latter. We "get to the point" only toward the end. However, I see no way to avoid the long introduction. In reality it is far too short, for it attempts a wide integrative or interpretative view based on many sciences.

I. The Domain of Information Science

C has such a broad domain, regulation and control, that it subtends all human activity. However, other fields such as physics, biology, and the social and behavioural sciences are equally broad. None are independent, nor possible to isolate in the universe of knowledge. Therefore to "define" IS can mean, realistically, no more than that it too can present certain aspects of this universe in useful, distinct cross-sectional view.

Since we aim at basic concepts, let us keep their number small. In C we use two concepts: variety, and the law of requisite variety. The latter is described later, so we begin with variety. This concept has gained a central place not only in C but in biology, psychology, and other fields. As will be seen, it is a broad "integrating concept". Variety is a property of a set, not of an individual. It is simply the number of discriminable differences which an observer can make in observing some system. Since the discrimination is made by the observer, the variety in a set may be more (or less) for one observer (or discriminating system) than for another. Variety is measured either as the number N of distinct discriminations, or as the logarithm of this number, $\log_A N$, where A is some arbitrary base (usually 2). The variety in the following letter:

Dear Dad: Please send money. Love. Your son,

is

1 (or $\log_2 1 = 0$) if the set is the message <u>as a whole</u>, or the unit set;

8 (or $\log_2 8 = 3$) for the set of words in the letter;

12 (or $\log_2 12 = 3.58$) for the set of letters in the message,

assuming we all count and discriminate identically.

This concept will grow in usefulness as we go. We now introduce at greater length some proposed basic concepts for IS. These are: 3-segmented IS path, SD and LD modes of message propagation; and later certain embryonic models such as a minimum abstraction tree model, and a rate model for information search. Some of this has appeared.[2-7]

Figure 1 is a diagram representing schematically a one-way path of message propagation from an individual A to an individual B:

The over-all path consists of three segments, each composed of many stages. Segment *ab* is in individual *A*, the author or sender of the message. Segment *bc* is in the environment or medium [or media] surrounding *A* and *B*. Segment *cd* is in individual *B*, the recipient of the message. The entire path is physical in the sense that all of its stages are subject to the laws of physics. We will call *bc* the external path segment (also "external segment," "external path," or "physical link"), *ab* the organic efferent path segment, *cd* the organic afferent path segment. Individuals *A* and *B* possess many peripheral sense organs and motor organs, but these are represented in each by a single effector or motor organ *M*, and by a single sensor or sensory organ *S*. The organ *M* is used to modulate suitable physical systems used as message carriers through intervening medium *bc*. The organ *S* is used for reception and transduction of the modulation within the same sense channel. However, it is not necessary that both *M* and *S* correspond to the same sense channel, provided suitable translaters and transducers exist. Each individual, *A* or *B*, is provided with a nervous system that conducts afferent modulation from peripheral sense organ *S* to central region *C*, which we call simply the "mind." It also conducts efferent modulation from *C* or a region near *C* to peripheral motor organ *M*. The exact locations of *M*, *S*, and *C* within the body boundaries are not material to our picture. But the fact that part of the path of propagation lies in each body is essential.

It will be noticed that what the communication engineer usually thinks of as the path of propagation is external path *bc*. The stages of this part of the path extend from the boundary of the message sender through possibly many media to the boundary of the message recipient. These stages may have widely differing conditions of propagation. They may include natural media such as air, water, and solids; or man-made media such as transmitters, receivers, and information-storage and -retrieval systems. *bc* may include organic, possibly living structures. But in general, the media outside the boundaries of *A* and *B* are purely physical and "non-semantic," in the sense that they do not contain the special message-initiating and message-receiving equipment located in communicating organisms. Thus, relative to *A* and *B*, *bc* is simply a set of stages not including the bodies of *A* and *B*, through which the modulation passes without the special processing that occurs in nervous systems (later referred to as "association"). In *bc*, modulation remains invariant if it is propagated in the absence of noise. We assume the range of propagation to be such that the power level remains high enough, and distortion low enough, for complete discrimination by *B* of the modulation encoded by *A*. The message may be amplified, regenerated, and transduced many times. But it remains simply modulation, physically transducible into itself in the form in which it left *A*'s modulating organ. With noise, the modulation deteriorates according to the second law of thermodynamics, and its "bit content" of information deteriorates according to the mathematical theory of communication. In many ways the purely physical modulated carrier in path segment *bc* is the simplest form assumed by a message as it passes from *A* to *B*.

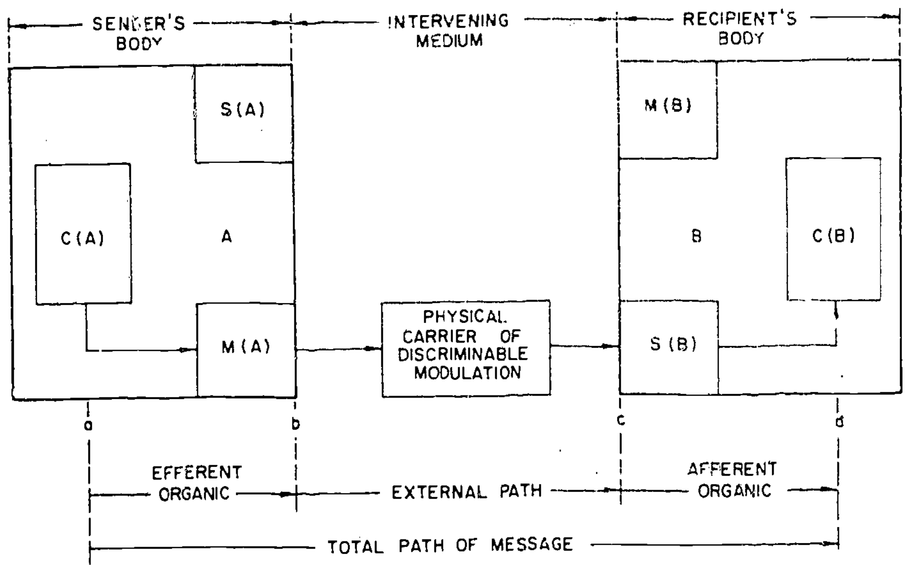
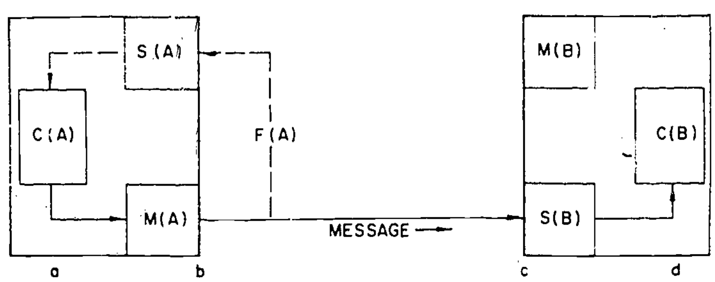
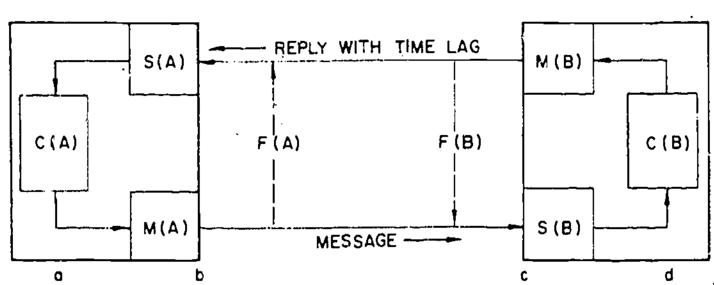


FIG. 1 Path of propagation of a human message.



(a) ONE-WAY COMMUNICATION
"MODULATION RECTIFIER" M(A)S(B) — COUPLED
WITH M(A)S(A) FEEDBACK CONTROL



(b) TWO-WAY COMMUNICATION
TWO MODULATION RECTIFIERS COUPLED
M(A)S(B) AND M(B)S(A) WITH DUAL CONTROL

FIG. 2 Simplest units of human communication.

Our first task will be to sketch a theory showing what communication is— how a message originates in a biopsychological medium *ab*, passes through a purely physical medium *bc*, re-enters a biopsychological medium *cd*, and conveys meaning from mental terminal $C(A)$ to terminal $C(B)$.

One obvious fact is that the stages in segments *ab* and *cd* are the only structural invariants in the path. The organic structures in *ab* and *cd* are built-in, fixed. Neural messages travel more or less fixed paths, paths (with almost no exceptions) beyond the power of either communicant to alter. By contrast the external path *bc* or "physical link" is almost infinitely variable by man. Because of this, what occurs in organic segments *ab* and *cd* is more characteristic of the message than what occurs in segment *bc*. The terminal segments are the only parts of the path in which meaning is encoded and decoded.

A second point is that the human one-way communication path resembles a simple control circuit. In Fig. 2(a), as in Fig. 1, a message is being sent by A to B. When the modulated signal leaves $M(A)$, A has little control over it unless he monitors what he sends. If he is speaking, he obtains feedback through $S(A)$, and adjusts his voice. If he is writing, he controls his message by visual feedback. This feedback control of the modulated efferent signal is shown at $F(A)$.

A third point is that in operation the human one-way communication path resembles a rectifier. If B decodes A's message and encodes a reply, the reply does not return through B by the same path as that of the incident message. The message to B travels afferent path $S(B)C(B)$. The message from B travels efferent path $C(B)M(B)$, with external return feedback through $F(B)$ [Fig. 2(b)]. For one-way mind-to-mind communication an afferent path in B must be coupled to an efferent path in A, across the medium. For two-way communication the coupling is both $M(A)S(B)$ and $M(B)S(A)$. A series of messages and replies (or conversation) produces an intermittent unidirectional "current" of modulation in a closed path. Terminals $C(A)$ and $C(B)$ function alternately as modulation "generators" and "recorders," with a time lag or lead. Thus human communication resembles vacuum tubes and solid-state devices, in that it exhibits a rectifying action. The complete path *ad*, the basic unit of human one-way communication when two persons are *M-S* coupled, might be called an "information rectifier" or a "modulation rectifier" [Fig. 2(a)].

When two paths *ad* and *da* are coupled for two-way communication [Fig. 2(b)], the two modulation rectifiers coupled in series (*M-S* and *M-S*) form the simplest complete system of human communication. As the message is encoded, sensory feedback at $F(A)$ and $F(B)$ monitors its alternating physical transduction. The message itself acts as a higher or "semantic" control feedback, altering and controlling the meaning encoded in each reply.[2]

Before generalizing this model, let us sketch in one more feature:

All messages may be divided into two classes: those of short duration (SD) and those of long duration (LD). Of these, SD are the more basic in the sense that we could communicate with only SD messages but not with only LD messages. A message in its simplest form consists of two components. The first is some physical system, which we will call a "carrier," that is not in itself a message (examples: radio waves without the voice or music "intelligence"; a blank sheet of paper). The second consists in descriminable marks on carrier such as images or sounds (SD) or printed letters or drawings (LD) which we will call "modulation".[2] The carrier can exist without modulation but not the reverse. In SD messages, the modulation varies in time. The marks on sound waves in direct speech change constantly at the ear - in fact they must die away (be attenuated) rapidly, if we are to be able to discriminate the next words or musical notes. If they persisted for even a short time more than they do, the sounds of successive speech or music symbols would become indistinct and blurred. Reverberation would destroy the meaning conveyed by the modulation. It is of the essence that SD messages be attenuated at least as fast as they pass into the sensor of a human recipient or of a machine-receiver. Somewhat more formally: the attenuation rate of the channel which conveys information to the sensor must equal or exceed the information rate of the sensor. By information rate is meant the time rate of change of fully discriminable "least units" of information such as word-symbols, or of their components, such as "bits". Unless we refer specifically to bits/second or other rate units, by "sensing rate" we shall mean "words/second".

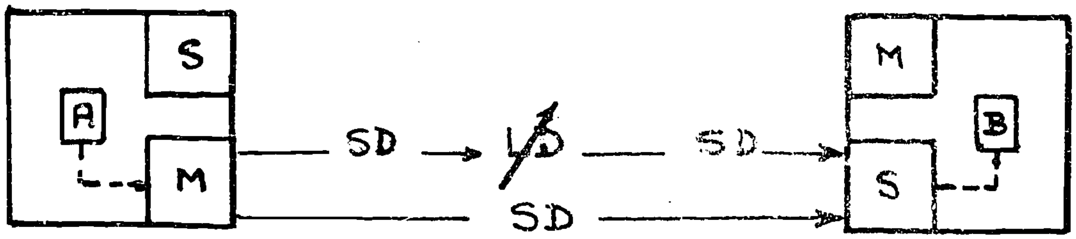
The reason why SD messages are more basic than LD is simply that when the message passes into the sensor of man or machine it must do so in SD form. Human sensory (afferent) and motor (efferent) messages travel by time-varying modulation to and from their destination or source - usually the brain. The same is true of machines which pass information through a sensor into some "decision" mechanism.

In contrast with that of SD messages, the modulation of LD messages persists for comparatively enormous time intervals. In order to achieve this extension into the time dimension, the carrier is restricted in most cases to a solid, and the modulation, instead of temporal, is spatial. Printed letters on a page store their contained message for long periods. They do so by extending

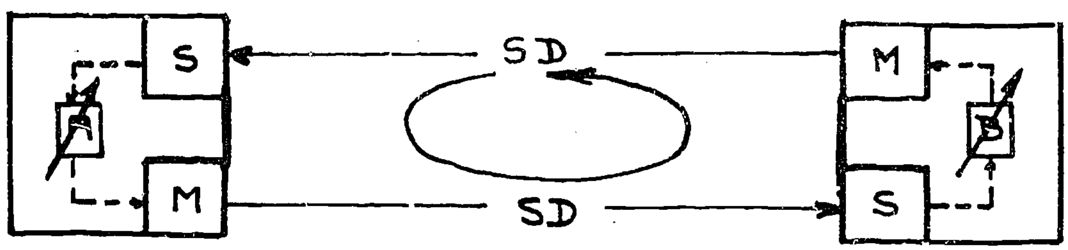
spatially on the page. Naturally, since both SD and LD modulation exist in space and time, both are four-dimensional "marks." But the far shorter time duration of SD modulation makes us refer to it as "purely temporal," which it is not; and to LD as "purely spatial," which it is not.

Because of this constraint on communication - that a message must enter the sensor in the SD mode - all "stored" or LD messages must be convertible to SD form. This is indeed the case. It is also true that many SD messages are convertible to LD mode, but this was not always so - and the conversion is man's peculiar discovery. He found that information in LD mode can be stored, i.e., propagated into the time[4] dimension, even beyond the life of the message sender. The discovery enabled the cumulation of knowledge - the possibility for 7
man's finite brain to tap a much larger memory than his own. Using this technique of storage as a tool, he erected science and civilization. But it is a basic constraint on use of the tool[3] that stored messages be retransformed into the SD mode.

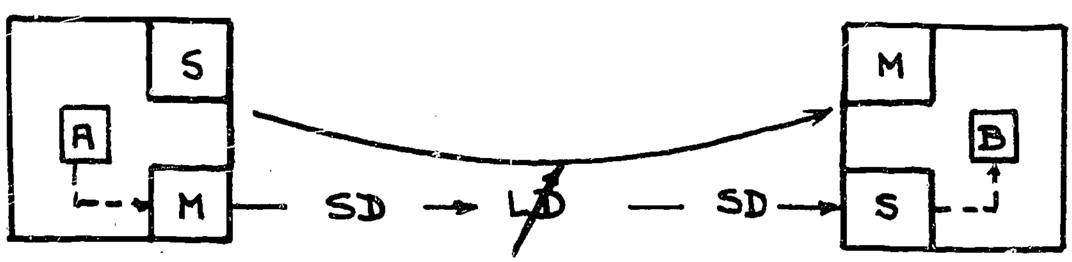
The reason for introducing SD and LD modes is that they enable us to define with some precision the domain of information science, as distinct from that of other sciences. Like all definitions, this one is only as sharp as the concepts (or the classes) used. Consider Figure 1 as an overall model of three coupled segments, such that the couping between segments is always in the same relative spatial and temporal order shown. Spatially, it does not matter where recipient B is located, but B is always separated from A by an "isolating" segment, bc. Temporally, A sends the message prior to the time at which B receives it. Two general classes of human communication now arise, depending on whether the message is propagated solely in SD mode, or is transformed into LD mode at some stage in segment bc, and then retransformed to SD mode for terminal coupling with B. This is shown schematically in Fig.3. With SD-only propagation the main time delays tend to occur at the terminals (source and sink, C; sender and recipient, B). This involves a relatively tight bond. The recipient usually must identify the sender, or at least have a channel that connects back to him. The path becomes a closed loop typical of the conversational mode, and also of feedback regulation and control. The closed loop predominates in all society, primitive and modern. Radio, telephone, television and other rapid signal transmission systems tend to retain the small time delays in the central segment, which are essential for closed loop communication. This is the main domain of cybernetics, for without control of the sink by the source possibility of regulation is greatly reduced. On the other hand open loop communication is the principal domain of information science. We now define it



A. BASIC 1-WAY MESSAGE PROPAGATION MODES:
SD
SD— LD—SD



B. SD: USUALLY PERMITS FEEDBACK.
CLOSED LOOP (CONTROL) COMMUNICATION



C SD-LD-SD: USUALLY BLOCKS FEEDBACK.
OPEN LOOP COMMUNICATION

FIG. 3 EFFECT OF MODE OF PROPAGATION
ON COUPLING AND CONTROL BY SOURCE.
: VARIABLE TIME DELAY

roughly, in terms of the concepts introduced so far. The proposed domain of information science is simply the set of all three-segment message paths (i.e., initial and final segments are living humans whose signal sending and receiving organs are appropriately matched to the channel of the signal in the central segment, the environment; and sender-environment coupling occurs prior to environment-recipient coupling) such that the mode of signal propagation is usually but not exclusively SD-LD-SD (or briefly, "SLS")

It will be noticed that, since SLS predominates on the "IS path", time intervals between the two couplings of the three segments may be very long. In fact, transduction at b may precede that at c by years, centuries, millenia. The LD message acts as a variable storage device and variable time delay which combines enormous versatility. A third kind of versatility is the "serial addition" of recipients, since the message may be "non-destructively" read out of its LD carrier at various times; and the "multiplication" of recipients through message replication. Thus the LD message enhances man's capacity to carry a message in his mind, lengthens its retention often beyond his lifetime, and extends the number of recipients beyond his capacity to contact them in space and time. If we regard the general goal of a message as "reaching a set of recipients", then the LD mode greatly increases the alternatives open to the sender for goal-fulfillment. The cybernetic "law of requisite variety" is as follows.

Let E represent a set of "essential variables" which must be kept within certain limits in order to achieve a goal. Let D be a set of disturbances or threats to E which, by acting on E through the environment T, can drive E's variables out of the region of stability for attaining the goal. Finally, let R be a regulator interposed to keep D from disturbing E by driving its essential variables out of bounds. Then the law states that R can successfully "regulate" D only if it has "requisite variety". That is, R must have a variety of alternatives sufficient to counter the variety of disruptive alternatives open to D. In stating and presenting this central law of cybernetics Ashby[8] represents the contest between R and D as a two person "matrix game". The elements of the matrix are the outcomes of the interaction of D and R on E, acting through environment T. The game consists in two moves. D plays first, by selecting a row of the matrix. R counters, by selecting a column. If the outcome (value of the matrix element jointly selected by D and R) is not within the essential set of E (E's region of stability for achieving its goal), then R wins by "regulating" D. Otherwise R loses. Only variety in R can "drive down" or "destroy" the variety in D.

By use of LD messages in addition to the basic and more primitive SD, man has enormously increased his capability to "regulate", i.e., reach, sets of recipients. The LD message, a

cornerstone of information science, may also be regarded as a cybernetic control device. It offers man increased variety of alternatives. The sender has the opportunity to "trade" comparative certainty of designation of a small set of recipients (usually known to him) using a closed loop, for uncertainty of designation of a vastly larger set of (usually unknown) recipients (open loop). Conversely, it offers the recipient a "trade" between comparative certainty of access to the messages of a small set of senders for uncertainty of access to those of a vastly larger set of senders. Viewed from IS the LD message is a space-time or four-dimensional switch; viewed from C it is a regulator against the "disturbance" of communication, with much more "requisite variety" for insuring that knowledge continue to be communicated than has the SD message alone. The messages of the arts and sciences represent communication with large sets of undesignated or loosely designated recipients. For them the LD mode, statistical from the sender's viewpoint, is universal.[9]

Finally, it will be noticed that the domain of information science has been defined in terms of phenomena accessible to social (objective) observation in middle segment bc; and of phenomena not as yet accessible, in segments ab and cd. We must be careful, however, to avoid suggesting that the signal, a physical pattern of free energy (energy differences) as it crosses segment bc, is ever observable in bc. It is not -- what occurs in bc must, like the messages that originate in ab, be observed in, and only in segment cd. Another possible misconception is that this "purely" physical signal "carries meaning" along with the pattern of energy differences which the sensor discriminates. Meaning exists as such only within humans. The locus of meaning in the universe is a set of small regions like those at the ends of the "IS path". Physical continuity of meaning in crossing gap bc of this path is an illusion. If a meaning seems to be "transmitted" from sender to recipient, what actually occurs is that the signal effects some change in the recipient which he interprets as meaningful. The change occurs in the interaction of the pattern of sensory variety with something else, considered below. The act of meaningful communication can be regarded from the viewpoint of IS as use of a signalling device whereby a sender operates switches in the mind of the recipient. From the viewpoint of C it can be regarded as another application of the law of requisite variety.

The essential discontinuiety between sender and recipient (the environment that intervenes physically between them and is represented by segment bc) introduces threatening variety -- variety which disturbs the basic need of individuals to act together, to form society. The variety in the ways of keeping the sender and recipient incommunicado is reduced and overcome by "taking advantage of the constraints"[10] These are, that it is physically possible to

propagate, not an entire sensory pattern, but a "homomorphism" of it, from one brain to another. This important mathematical and cybernetic concept has briefly been described in relation to IS[2]. We now try to follow the sensory pattern, beginning with the direct sense impression (DSI) produced while a stimulus is acting, and the stored sense impression (SSI) or trace left in memory after the stimulus ceases, and beyond.

II. An IS Theory of Communication

We begin with a paradox. In segment bc what is propagated is a physical pattern, "modulation".[2] All our cognitive mental contents either consist in such patterns or are derived from them. We can divide these mental contents into those that we cannot directly communicate (DSI's and SSI's) and those we can (something else). The paradox is that the signals that cause our DSI's travel, while, given a DSI, we cannot originate a signal that will convey that DSI directly into another mind. I cannot transfer into your mind the scene I see. The paradox disappears if we recall the rectifier-like nature of the path of a 1-way message. The pattern of your DSI must originate outside you, in segment bc. But it cannot originate farther back, in my segment ab. However, since we do communicate <u>something</u> all the way from a to d, then that something cannot be a pattern of one of my DSI's. RATHER, it is a <u>subpattern</u>, derived from DSI's and SSI's that entered me but which I cannot transmit out again. In other words, we cannot communicate the full patterns present in DSI's and SSI's, but only partial patterns extracted from them.

A huge literature exists on the subject. It is not known, however, exactly how the brain extracts subpatterns from the full patterns. It is known that, as the DSI's are stored as SSI's, the brain <u>associates</u> them, possibly by the growth of bonds at the sites of the memory traces. For purposes of information science we distinguish two main kinds of association, "natural" and "conditioned". Natural association is less studied in psychology than conditioned. It is less controllable. It is responsible for the spontaneous (without our conscious control) formation of our concepts of objects, for our "reconstruction" of the environment. Conditioned (Pavlovian) association can be under conscious control. It occurs when we associate an arbitrary pattern (for example, a visual word pattern, or auditory sound pattern) with a natural pattern. That is, we "attach" the arbitrary pattern (symbol) to the naturally formed pattern (concept). Since the bond of conditioned association to a natural pattern can be made as tight as we choose, then the formation of the symbol as a DSI in the mind of a recipient will actuate or evoke the natural pattern into his awareness. (It can also be recalled or evoked by a DSI which originally formed and is part of the natural pattern, e.g., by suddenly seeing a scene, or hearing a sound.) In this

way the sender of a message can send down his path (ab) a partial pattern, namely, that subpattern of a DSI necessary to form a symbol. The symbol is transduced at b, into the apparently full pattern which eventually reaches the recipient at d, and evokes a concept. The sender of a symbolic partial-pattern "plays" on the set of stored concepts in the recipient's mind, evoking one after the other, and in the process causes the recipient to experience a message.

This sketches the process of symbolic connection -- the use of arbitrary but socially accepted symbolic patterns to evoke the natural ones which our minds form spontaneously. The real mystery remains: how the brain forms concepts or partial patterns, out of complete sensory patterns in the SSI's.

III. Pattern Recognition and Abstraction

It is probable that every DSI remains "nothing but" a physical pattern until, perhaps, it reaches the cortex. In other words, there is a stage before which a DSI is merely a pattern, and after which it is "recognized", and in so doing "acquires" meaning. The process of recognition is preceded by another, called "feature extraction". It is analogous to what Newton found occurs when a beam of white light is passed through a prism. A parallel beam of white light is decomposed into diverging colored beams. If these colored beams are collected and passed back in one direction, they recombine into white light. That is, a stimulus (beam) which produces in us the perception of "white" light is analysed by physical structure (prism) into components that we then perceive as "colored" light. Analogously, when a DSI is still "merely a pattern" on the cortex, it contains, within its complete pattern of discriminable differences, simpler subpatterns into which it can be decomposed. A subpattern of a fuller pattern is a "feature", or "characteristic" or "property" or "internal relation" of the fuller pattern. To perceive the feature requires simplifying the full pattern-- merging or erasing all but that which displays the feature. Now since the extracted pattern is no longer a full sensory display (the "concrete" kind which we associate with "reality") the extracted subpattern will not appear to us like a concrete DSI. It will contain one or more extracted features or relations. We call it an "abstraction". It is well named. It is literally abstracted from sets of SSI's which contain in their patterns the subpattern or feature. It is less than fully pictorial, and more relational, as indeed are our abstract concepts. That is, it does not contain more relations than the original DSI, but more relations relative to its reduced content. The brain receives and stores perhaps millions of SSI's. From them are formed, in the infant rapidly and adult more slowly, the partial-pattern biophysical structures that will, once formed, decompose new DSI's into abstract components, much like white light into colored

components. Learning or induction consists in establishing the abstraction-structures. Deduction consists in using established abstractions to interpret new DSI's (experience), and predict their nature from their spectrum. When a new DSI pattern interacts with a set of partial-pattern structures it is rapidly and reliably decomposed into the abstraction spectrum of its properties. This enables us to instantly understand something we perceive, and respond "intelligently". And only in terms of this spectrum is a DSI either understandable or describable with symbols. The spectrum of its properties comprise its meaning. In this way a DSI, a pattern, up to some stage without meaning, through decomposition into partial patterns instantly "acquires" a characteristic spectrum which identifies it and is its meaning. The response time, especially if the abstraction-structures are well established, is very short, and we are unaware of an interval between "meaninglessness" and "meaningfulness". At present a whole new field of psychology is opening around the complexity and size of the abstraction structures as measured by response time. We are of course unaware of all the many interactions going on. We perceive the spectrum, not as distinct relations but as a composite of properties -- the "meaning". Actually a meaning is a set of partial-patterns or concepts. Philosophers have long been aware that "an object" is intellectually equivalent to a set of properties observed through its behaviour.

A great deal more could be said about this process, central to mental activity. Evidently it bears on human development -- the rate and kind of partial patterns that are set up in the brain bound the meaning which at any time the person can attribute to any DSI experienced. It bears on education -- the more or less socially planned, guided or aided establishment of partial patterns. As will be seen later, it determines what the librarian and information scientist does in "organization of knowledge". But before discussing this it is necessary to review several other matters.

The point has been made that only by means of the abstract properties into which it is mentally decomposed can a DSI acquire meaning. The same is true when we try to communicate it by composing its meaning in another mind, or "describing" it. Since we cannot project the DSI or any concept directly into the mind of another, we string together some of its most salient properties (to us) and evoke them in the recipient's mind with a corresponding string of symbols. But this is the inverse of decomposition. In our own and the recipient's mind it is composition. In description we use symbols to reconstruct our own concept within the mind of another, much as we reconstruct white light from previously decomposed colored components. Providing the recipient also has the same set of abstractions, our symbols

evoke them in him and he "follows" rapidly and easily. Indeed, if he happens to have a more detailed or finer set of abstractions, he may by association perceive more meaning (subjective to him) than does the sender. Any description, any communication of meaning, requires at least a few shared abstractions, and is possible only because their partial patterns have been formed in both communicants. So in communication we invert the analytical process of feature extraction, and synthesize the described concept. One of the prime problems of DESCRIPTION IN science is ~~[illegible]~~ HOW TO INSURE that in reconstituting any concept through its spectrum, the natural order of DSI's observed in a phenomenon is faithfully preserved in the concept transmitted by symbols. "Operational" definitions attempt to retain "objectivity" in scientific description by describing concepts only in terms of those derived from actually or potentially performable operations.

"The concept is synonymous with the corresponding set of operations". [11]

Without such safeguards it is easy to synthesize in the mind an apparently plausible but actually inconsistent mental construct. The assumption that nature cannot be inconsistent -- only those who describe nature -- is the basis for much of scientific method, philosophy and logic. We now consider another aspect of abstraction.

IV. A Theory of Classes and a Simple Model of Minimum Class Structure

We have discriminated between natural association which gives rise to our concepts and conditioned association which gives rise to the arbitrary sets of symbols which constitute languages, and by means of which we evoke or trigger our concepts from their latent state in memory to temporary activity in our awareness. The role of symbols has been much more fully explored than has the role of natural associations or concepts. The relations abstracted from our SSI's correspond to "invariants" or "constraints" in the environment. The environment stamps our DSI's with patterns containing these relations, and our mind abstracts them as its properties. Thus all natural laws are constraints discovered among our I-sets. More formally, they are homomorphisms. Homomorphisms have previously been considered purely mathematical concepts, until they were generalized by cyberneticists.[12] Homomorphisms are relations between structures. They are discovered when we observe that one structure (or system, which may be real or purely mental) is similar to a simpler structure, but only if we "map" the more complex structure onto the simpler one in a certain two-step way. The first step is to simplify the more complex structure by "merging" or ignoring some of its complexity, i.e., suppressing some of its structural variety. We need not actually lose the discrimination of the variety. We merely ignore it for purposes of simplification. Merging or ignoring means to treat

as if equivalent. Since variety consists in discriminable differences destroying variety means reducing it to a state of non-discrimination -- to equivalence. The more complex structure now has fewer "parts".

The second step is to map the reduced complex structure onto the originally simpler structure, by a 1-1 correspondence. If this is possible, the simplified structure corresponds, part for part and structural relation for relation, to the parts and relations of the second, originally simpler structure. The two structures have become isomorphic. They are behaviourally indistinguishable. Isomorphism does not mean that the two sets of parts and relations are identical. It means that, viewed abstractly, they behave or function alike. For example, an electric circuit or a computer can be made to function isomorphically with a mechanical system or machine. They undergo changes synchronously, but the nature of their responses are electric and electromechanical and mechanical respectively.[12] Likewise, the abstract concepts into which the brain analyses or decomposes DSI's consist in simplified patterns, homomorphisms abstracted from the more complex sets of SSI's, or I-sets. We experience the abstract patterns as properties of a DSI, or of any object observed; or in fact, of any mental object to which we can ascribe properties. As an example, there is a homomorphism between the complex concept "spoken and written language without metrical structure"[13] and the simplified concept "prose"; and an isomorphism between the discovery that one has been speaking prose for more than forty years,[14] and the discovery that the concepts of prose (and poetry) are homomorphisms. An abstraction is a homomorphism -- a simplification of a more complex structure, and a 1-1 remapping onto a simpler one. This process destroys some of the variety in the original structure, but conserves some of its internal relations. [redacted] Symbols associated conditionally with the homomorphisms MAY tend to reflect some of the conceptual homomorphic structure, but because of the arbitrary nature of the symbols and the partly arbitrary nature of their connecting syntax, we should not expect to find linguistic structures accurately nor uniquely mapped on conceptual structures. The same conceptual mapping may be performed by different [redacted] LINGUISTIC STRUCTURES in a given language (the dictionary phenomenon); and in different languages. In each mapping the final isomorphism is identical with some simplified concept. But are the differently derived isomorphisms equivalent, even in the same language? We must expect some difference, and where it is great enough, conscious discrimination in the form of "near synonyms". As between languages the question of whether a unique isomorphism of conceptual structure can ever be communicated (that is, so as to be independent of language) concerns the linguistic study of translation, the attempt to preserve conceptual invariance in remapping with more widely different symbol sets.

The basic assumption underlying the above ideas is that the brain, starting with sensory DSI's, abstracts from them qualities, characteristics, features, relations, abstractions, abstract concepts, or categories and classes. It is this last aspect that we pursue.

According to Piaget understanding arises as the child develops the categories or classes into which experience is analysed.[15] The same conclusion was reached by Bruner, Goodnow and Austin:[16]

> Much of our commerce with the environment involves dealing with classes of things rather than with unique events and objects. Indeed, the case may be made that all cognitive activity depends upon a prior placing of events in terms of their category relationship. A category is, simply, a range of discriminably different events that are treated "as if" equivalent.

The repetition of these ideas is more apparent than real. The aspect that is new is that properties also determine classes. Thus "features" are decision rules, "intension" in logic, that "classify" DSI's, determine their class memberships in abstract classes. The number of DSI's that conform to such a rule are the class "extension". This inseparable mixture of qualitative property which constitutes a means for deciding "like or unlike" in property or pattern, and quantitative property which designates the number of instances, particulars, cases, or members, occurs in all abstract thinking. Every concept, no matter how abstract, has some defining intension or decision device, and a membership or extension. The DSI has the maximum intension we can crowd into one experience and, of course, an extension of one. That it is also unique, is another aspect of unit extension. Quoting, (READ SSI FOR DSI HERE AND IN FIG. 4):

> Restated, the theory is that the qualitative property or properties of a set of associated DSI's which constitutes an abstraction from the set is the intension that defines a class. And the extension of the class is the number of DSI's associated through the abstraction class. This may be a rather large number, but is always finite.

There are a great many mental classes, some of which overlap in extension through interlinked DSI's. Many more overlap through interlinked abstractions. The number of distinguishable characteristics of "units" of intension (variety) abstracted from a DSI or set of DSI's varies greatly. We experience this as abstractions more "concrete" (more like DSI's) and more "abstract" (more removed from DSI's). "Image" is more abstract than "color," "color" is more abstract than "green". This relation of ascending abstraction can be demonstrated in classes that have a special

relation-class inclusion. They exhibit what are sometimes called "levels" of abstraction. Consider the sequence shown in Table 1. It is not suggested that the differences in intension shown in the table are the minimal steps. Introspection frequently shows intermediate steps where at first there appear none. The differences of increments in intension are suggestive only, with the exception of the first two. These are fairly certain. As the level of abstraction increases, the amount of detail or "pictorial" content decreases. Between levels 0 and 1 there is a large decrease, for even a "photographic" memory cannot retain the detail in a DSI. As the level of abstraction increases the minimum number of DSI's required increases. "Minimal" is important, for at each level above the second, the probable number of associated DSI's is greater than the minimum; at higher levels, much greater. As we go up, the minimum extension increases as 2^{k-1}, where k is the level (k=1,2, ...,N), 0 is the DSI level, and N is the highest level of abstraction derived from a given DSI or set of DSI's. More precisely, if we express the minimum extension in units of DSI's (which may be considered as roughly equal units of maximum intension, or as large units of sensory "bits") we have

$$I = I_0 - I_k = \text{intension}$$

$$E = 2^{k-1} = \text{extension} \quad (k = 1,2,\ldots N),$$

where I_0 is the maximum intension, namely the intension in the unit, or DSI. This relation involving a decrease from maximum intension, and growth of mimimum extension, is suggested in Fig. 4.

Fig. 4: Suggestive Model of Levels of Abstraction

TABLE 1

CLASS HIERARCHY SHOWING SUGGESTED LEVEL STRUCTURE AND MINIMUM EXTENSION, IN UNITS OF DSI's

Level	Mental object	Example	Minimum extension in units of DSI's and stored DSI's	
0	DSI	This hand and background, while you look at it	1	
1	stored DSI	This hand and background, as remembered (not immediately)	1	2^0
2	identity of a DSI	This hand and background while you look at it, and compare it with a stored DSI of same.	2	2^1
3	identity of a unit class	This hand abstracted from background, from two identified DSI's (may be possible with one DSI and one identified DSI)	4 (3)	2^2 (3)
4	identity of a class of unit classes	This identified left hand, and that identified left hand	8	2^3
5	identity of a class of a-class of units	Class of left hands	16	2^4
6	identity, 6th level	Class of left hands, and class of right hands	32	2^5
7	identity, 7th level	Class of hands	64	2^6
⋮ N	identity, Nth level			2^{N-1}

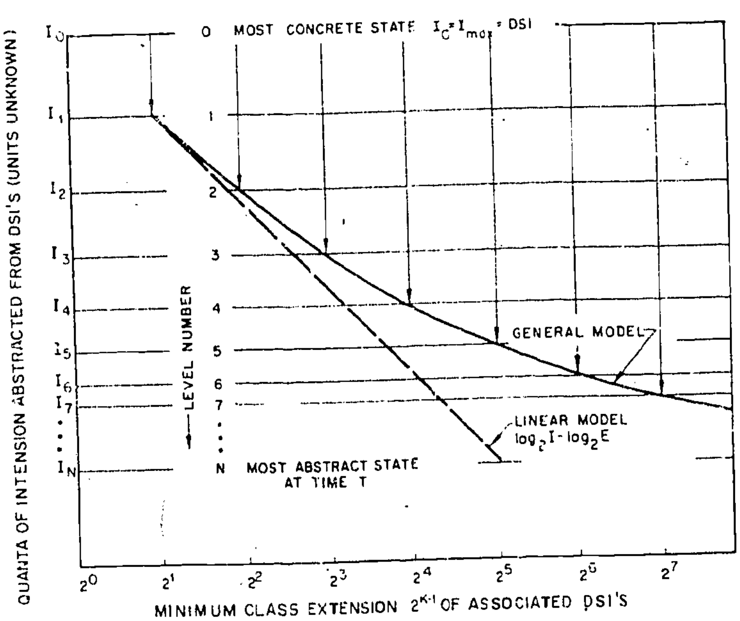


FIG. 4 Schematic hierarchy of classes abstracted from associated set of direct sense impressions at "ground" level I_0.

The significance of such a structure can be interpreted cybernetically. An individual DSI is unique. Yet the DSI, the only object that effects communication, cannot be experienced by two persons, nor repeated by the same pers[illegible] ness of experience [illegible] basis for sharing experience. [illegible] DSI's and SSI's [illegible] our only [illegible] mental contents nothing would be repetitive, nothing similar to any other thing. We could not compare experiences. Ashby states: "The most fundamental concept of cybernetics is that of 'difference', either that two things are recognizably different or that one thing has changed with time"[17] True, but psychologically, "equivalence" is even more fundamental. Just as one cannot discover invariance without change, one cannot experience change itself unless something remains constant. Therefore, starting with DSI's and SSI's only, we should possess so much variety that we could not experience or conceive of "difference". The significance of the new structure is that it destroys enough variety so that differences can be observed AND REMEMBERED. The senses supply ~~variety~~ but not ~~[illegible]~~ RECALLED differences. Constancy is necessary for becoming aware of differences. Sufficient "prerequisite constancy" is necessary for use of "prerequisite variety" What is sufficient is discussed in the next section.

Assuming that our understanding and knowledge (e.g., of natural laws) arise through constraints on sensory variety[18] and that the brain creates these constraints (or constancy) as a bio-psychological artifact, homomorphic transformation becomes the chief higher mental process. In homomorphic transformation some but not all variety in a complex structure is suppressed or merged. This does not mean that it is lost The SSI's can be recalled intact from memory. Their variety then must be suppressed in some added, semi-independent structure. In this structure is replicated (perhaps by a matching reminiscent of the transfer role of RNA relative to DNA) not all the variety in the SSI's but the set of reinforced partial patterns. The new auxiliary structure must simultaneously act as a semi-independent "equivalence class", and yet remain connected to the individual memory traces (SSI's) so that if one of them is activated (as in sudden recall of a past scene) the interpretive abstract classes are also activated. The whole functions according to the law of requisite variety. The "disturbance" D is the variety in the SSI's. They and their natural order of occurence are the environment T. The auxiliary (abstract) structures are the "regulators" R constructed and interposed by the body between D and the essential variables E among which are our survival-oriented, "intelligent" responses. The set of abstractions is a superb cybernetic device. It creates both the needed invariance to support observation of differences on which depend internal communication or thought, and "sufficient similarity" between minds for external communication. (see next section). It is a masterpiece of evolutionary achievement and simplicity. All the effects are accomplished in one economical process -- erasing sensory differences. The cognitive process can be visualized as two main cybernetic steps: data gathering, or production of variety through sensors; and data processing, or homomorphic simplification through natural associative pattern filters -- reduction of variety into abstractions. Each process and supporting brain structure requires the other. Together they constitute the basic devices of cognition.

V Requisite Constancy

We have conjectured the existence of structures in the brain (call them A-sets) derived from naturally associated sets of stored sense impressions (I-sets). Each member of an A-set has the nature of a relation, an equivalence class for its member SSI's and a qualitative "property" of the I-sets. It is also the intension which defines a class. For the special case of an A-set based on tree-like I-set we suggested a simple model, giving the minimum number of naturally-associated SSI's required to form the i^{th} "level":

$$n(i) = 2^{i-1} \qquad (i=1,2, ---m)$$

where m is the highest level constructed within the brain on that I-set as base. Then the total number of naturally associated SSI's would be at least

$$N(m) = \sum_{i=i}^{m} n(i) = 1+2+4+8+\cdots+2^{m-1}$$

$$= 2^m - 1$$

In this model each successive level contains one more naturally-associated SSI than the entire sum of all those in lower levels. What interpretation can be placed on this rapid increase in minimum class extension? Again there is a straightforward cybernetic interpretation: the levels represent <u>decreasing</u> variety or class intension (or partial-patterns) in the set of SSI's which define the class. A decrease in variety may be regarded in several ways. Most obviously, it represents a lessening of constraints, so that more objects can be found that comply with it as a defining rule for a class, i.e., class extension increases. Another aspect is that, if range in variety used to define a field is narrowed, the range in the corresponding concepts is narrowed, or the <u>concept stability</u> increases. That is, concept stability would be some inverse function of range in permissible variety. But class extension is such an inverse function (although not necessarily the correct function) Hence we may regard the increasing extension and decreasing intension in the abstractions as a measure (of some sort) of increasing <u>conceptual stability</u>. Finally, there is another aspect of the same phenomenon. The decrease in class intension also corresponds to increase in <u>versatility of response</u> behaviour, on the part of the person whose brain is involved. There are innumerable examples. For instance, when John is five, the question "how many are two cows and three horses?" is no poser, for he does not see the difficulty. As he matures, he senses that cows and horses differ and cannot be added. When he has developed still more levels, he calmly uses A HIGHER GENERIC ~~an upper~~ level, and answers that "two domestic animals and three domestic animals are five domestic animals." Thus he has acquired <u>additivity</u> by use of a higher, more abstract level. The

physicist does the same when he insists on "dimensional homogeniety" in the terms of his equations. The mathematician strips all the internal structure of intension from his mental objects, leaving only their externally discriminable structure -- their number and order. He thus creates the integers, and thereby gains still more versatility. Greater generality is equivalent to less defining intension of a class, to less variety. And the rule is that, the more general the class, the more "particular" cases it subsumes, and the more versatile is the mental owner in performing mental operations.

Now let us look in the opposite direction. At the bottom level is the SSI. It contains the maximum variety we can experience in one observation. It has the least versatility.Using only the SSI we cannot communicate internally in thought, or externally, in a message. If we look a few levels higher, we still cannot communicate. For example, we cannot communicate the level 2 in Figure 4. The uniqueness is still too great, i.e., there exist no comparable structures in the mind of another person. However, at a certain level the stability in conceptual structure in two minds begins to be <u>sufficiently similar</u> so that the indication of the concept structure by one person finds something similar which can respond, in the other person. When this hypothesized <u>threshold of stability</u> for communication is reached, there is an enormous simultaneous increase in versatility. For now the two persons can function as a social unit, mutually assisting each other's goals. This hypothetical level we indicate by n(c), the minimum number of naturally-associated SSI's for the threshold conceptual stability necessary for interpersonal communication. Evidence for the existence of such a threshold are the facts that we cannot communicate DSI's and SSI's, and that we can and do communicate by signs or symbols that evoke abstract classes. For levels higher than c, communication becomes easier and easier. The probability that our conditioned associations (symbols) evoke ~~abstract~~ concepts HELD IN COMMON (if they are present) becomes better and better. This idea underlies the ease of communication by small groups of professionals who share the same sets of abstractions. The precision of mathematical concepts, for example, is no accident. The fact that they can be (not necessarily always are) so precise is attributable to their enormous suppression of variety. They are actually derived from very large I-sets of which the mathematician loses awareness. In fact the difference between mathematics and other sciences lies in a kind of superversatility-- the mathematical concept structure is not necessarily reconstructed so as to embody the constraints of patterns of variety observed in the real world of DSI's. Yet one of these constraints persists in a way that the mathematician must observe. He uses it as his link with "empirical reality". The constraint on his synthetic patterns is that of validity of proof. Proofs allow enormous simplification (a homomorphic device) since they permit suppressing the variety in lower levels of abstraction, and retaining only the reduced patterns at the higher levels. A proof is a rule for interlevel transitions,

"inference", in logic. The example below applies only to levels of abstraction based on tree-like I-sets, as in the simplified model:

Hypothesis 1:	All A is B	(If A then B)
Hypothesis 2:	All B is C	(If B then C)
Conclusion :	All A is C	(If A then C)

Assuming the mathematician can demonstrate the truth of the conjunction of the hypotheses, then a valid conclusion follows, and the middle term B is unnecessary, as in fact is A, since the whole process is designed to show that A also conforms to rule C. This type of constraint, empirical in origin, has the versatility of applying to any three levels. The example merely suggests that logic too is a device for reducing variety, for regulating and controlling it, i.e., for increasing the versatility of its user.

Let us now see how these mental structures in IS path segments ab and cd determine physical access in segment bc. Intellectual access occurs, of course, only in segment cd.

VI Access to Knowledge As Control over Variety

To summarize our position: knowledge does not exist in the RECORDED literature [redacted] nor in any symbols for classes. Knowledge exists only at the terminals of the IS path, in the form of abstract classes organized from and connected with stored sense impressions. The literature through which access is obtained exists in collections of LD messages containing symbols. Therefore, in discussiong access to information and knowledge what we really mean is access to arbitrary patterns, located externally, of themselves without meaning. By prior conditioned association, however, they can, when sensed, interact with and evoke into awareness our internal, naturally associated patterns. The new internal superpattern temporarily reconstituted in awareness can, through its texture of choice and order, convey information and add knowledge. The abstract mental pattern is the basic form of reduced variety. The greater its stability the surer the contact between the external symbol and the delicate structures of understanding which only abide within the living organism.

The problem of access to knowledge can be analysed in various ways. One of the most systematic would be to take advantage of the structural organization provided by the IS path, and consider it by segments and stages within segments. Because the path of propagation of a message along this path is directed, a directed-graph representation might be useful. We should regard each stage as a channel, limited in its type and capacity to transmit quantity of variety per second. A model has been based on the rate at interface c (rate of sensing symbols, on the part of a human or a machine). This model showed that limited channel rates for sensing variety have shaped all IS systems:

"due to the accumulation of LD messages, any inquiry takes place among vastly more messages than can be sensed in the available time. Out of this constraint on use of stored information arose, we suggest, the main framework of the system of homomorphic transformations which underlie "bibliographic access" to stored literature."[3]

Let V be the average sensing rate for units of variety (e.g., words/second). Let T be the number of seconds during which sensing occurs. Then

$$VT = M$$

is the total number of words sensed during T, or the "message length", measured in words. This quantity is numerically equal to some fraction of the collection sensed:

$$VT = M = \frac{N}{CK} = \frac{nA}{CK}$$

where N is the total number of units of variety (words) in the collection, n the number of major units of variety (e.g., documents or volumes), A the average number of words per major unit (words/document, or words/volume); K is the average homomorphic compression factor, and C the initial selection factor. Since the main constraints are in the small magnitudes of V and T, (that is, we can sense only a short message within any reasonable time) we organize our searches around the size of the collection N, the compression K ($K = \frac{\text{number of words in original message}}{\text{number of words in compressed message}}$, a ratio, or pure number without dimensions), and the power of the classifying system to divide the collection, i.e., to eliminate all but a fraction N/C of the original collection (K=1), or N/CK of a compressed or surrogate collection. C is also dimensionless. The searches take place on the compressed collection, and their objective is not to sense the literature but to decide what literature to sense. Without going into details we may say that the ranges of C and K are the same:

$$1 \leq C \leq N \qquad\qquad 1 \leq K \leq N$$

and that the types of searches depend on the numbers of prior-prepared compressions (values of K), and structural organization which permits selection (dividing the collection by C). In all cases, except that of direct sensing of an original collection (K=1), the search starts at the highest value of K. N/K is the total compressed collection. If access is, for example, through catalog cards only, then the collection is compressed unit by unit (by volume), and K would be

$$K = \frac{\text{number of words per volume}}{\text{number of words per card}}$$

For example, if $A = 100{,}000$ words/volume, and the average number of words per catalog card were 100, then

$$K = 100{,}000/100 = 1000$$

For such a system the sensing takes place on the "image" or "surrogate" collection of cards. At the start of the selection $C = 1$ (nothing has been selected). After one word had been eliminated, the value of C would be

$$C = \frac{\text{Original number of words}}{\text{Remaining number of words}}$$

$$= \frac{\frac{N}{K}}{\frac{(N-1)}{K}} = \frac{N}{N-1}$$

in other words, during search C <u>increases</u> from 1 to N (the end of the search, when the last word is about to be sensed); while K has one or more constant values, which <u>decrease</u> until $K = 1$. In all cases the manipulation of K and C follows a certain order. Since the act of selecting requires variety of more than 1, the values of K must be reduced <u>before</u> the value of C can be increased. For example, suppose a system is used (somewhat like the Dewey Decimal System) in which the collection is first divided into ten parts. If the entire collection were first represented by one word, say the word "Collection", then the value of K would be, initially

$$K_o = \frac{N}{1} = N$$

and no selection could be made from the one word. If however, K were reduced so that (decreasing K by a factor of ten)

$$K_1 = \frac{N}{10}$$

then it would be possible to select one or more of the ten words. Thus the variety in the compressed collection had to be <u>increased</u> from 1 to 10, <u>before</u> the selection from a variety of 10 could be made. In any decision process, there must be a variety of at least 2. In information search with a classification system we alternate decompression (the inverse of the homomorphic compression that took place when the more abstract class was formed), with selection. We go from the abstract to the concrete, always keeping the amount of variety to be handled at each stage small (and therefore the sensed message M short). In this way the total sensed message:

$$M = VT = VT_1 + VT_2 + \dots VTm$$

$$= M_1 + M_2 + \dots\dots Mm$$

where m corresponds to the lowest value of K greater than 1. At $K=1$ the nature of the search changes from decision-making (selecting what to sense) performed on the homomorphically-mapped image-collection, to sensing the selected part N/C of the original collection. Advantage can be taken of this change in function, as described below. The successive reductions in K permit the selection process to take place rapidly (since at each stage j only the few words M_j are sensed) and still preserve exhaustive coverage, since the homomorphic compression represents the whole remaining collection at each stage. Figure 5 shows schematically the inverse changes in K and C, with those in K preceding those in C. Note that after $K = 1$, the rate of climb of C is much slower, since the "reduction" of the collection now occurs only at the ordinary reading rate V. The entire time of search can be symbolized by

$$VT = VT\,(K>1) + VT\,(K=1) = M\,(K>1) + M(K=1) = M$$

in which the length of M and T are controlled by regulating the amount of variety sensed. The analysis also suggests that the two major cognitive variables in access to knowledge are contributed by two types of homomorphic mapping: selection, in which the part that remains is isomorphic with the original part; and compression, in which the new structure is homomorphic with the entire original structure. We are thus led to the generalization that search for information in an LD collection consists in strategic use of PREPARED homomorphisms, combined with instantaneous isomorphic selections. It suggests that there is an approximate constant:

$$KC = N/VT$$

or rather, a parameter N/VT determined jointly by the size N of collection and VT, the message length convenient for the speed and time available to the sensor. This hyperbolic relation between K and C holds, of course, only over the region in which both vary, namely, the region of decision-making or selection, $K>1$. Since C can vary only with change in K, it further suggests that designers of future information retrieval systems, or planners of search strategies, should consider the number and values of built-in K levels as prime factors in design. K and C respectively reflect the two suggested basic mental processes: K, homomorphic simplification and suppression of variety; C, isomorphic sensing of variety, coupled with its elimination by selection. K and C can both be regarded as homomorphic mapping processes, one by compression into a simpler structure with retention of certain invariants (the meaning), C by mapping an original structure into a binary function (1,0). The result of the latter is to select parts of the original without otherwise altering them.

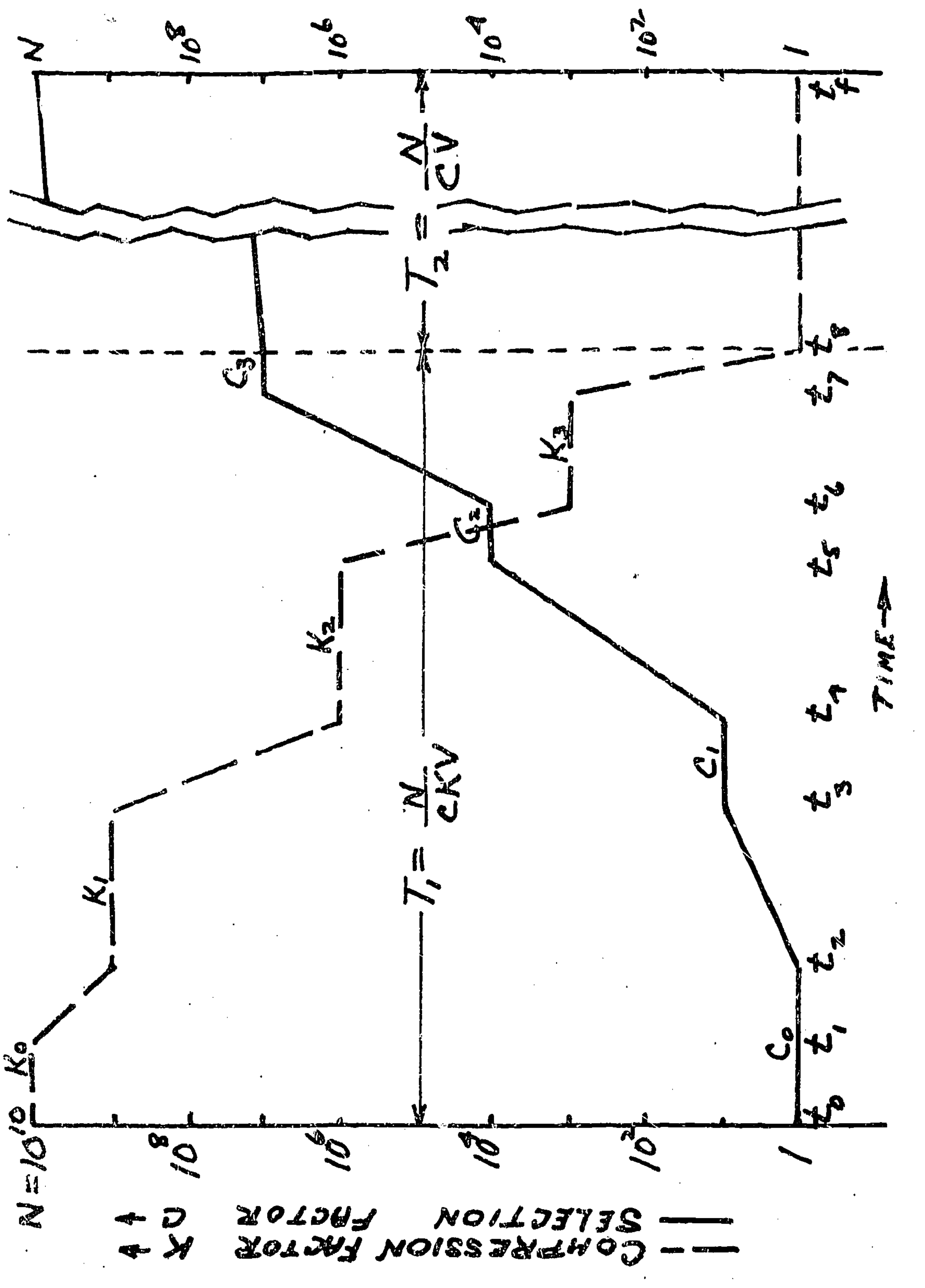


FIG 5 ALTERNATIONS OF K AND C BLOCK VARIETY IN HIERARCHIC SEARCH

Another suggestion for improving access arises from this model. Machines can now recognize patterns of variety that humans cannot; and for patterns which both can recognize, machines can be more rapid and reliable. A "trade" is possible between a wide range in variety of patterns recognized by humans (for example, in reading various kinds of handwriting and printed fonts) at slow speeds, and a much narrower range in variety necessary to make the machine reliable. The machine requires a degree of stability of symbolic pattern much higher than does the human. If this cost is paid, however, machines can sense at much higher rates than can humans. Since they must extract features from the patterns in order to recognize them, this "classification" by features also permits feature-class coordination. Thus machines can "make decisions" and "select", without understanding the meanings of the classes coordinated. This is one advantage which can be taken from the separation of the functions. The machine relieves the human of the work of sensing $VT(K>1)$, and leaves him only that of sensing $VT(K=1)$, where the meanings are required.

An extreme saving of another kind can occur if the machine pattern recognition speed is great enough. Suppose $K=C=1$, that is, there is no compression, and no selection. Then a pattern recognition machine that had a speed

$$V = N/T$$

could entirely despense with "classification". The decision-making preliminary phase would be eliminated, and the entire collection N could be "unstructured" yet searched in acceptable time T. This solution would not, of course, allow a human to derive meaning from the sensed collection. It would, however, allow the human to select from and locate within the collection any pattern or combination of patterns. For certain purposes the unstructured collection might serve as well as the classified collection. Prior to the age of machine pattern recognition humans had no choice but to classify collections. Thus the reduction in variety represented by the resulting machine capability gives increased versatility to the human "regulator" of the variety.

Compression of the range in variety underlies all physical compatibility of real systems of machines. Dramatic examples occur when, for example, the gage of a railroad track is standardized, the dimensions of camera films and microforms are agreed upon by the makers. They permit interchangeable use of railway cars on different lines, interchangeable cameras and developing equipment, uniform microform storage and viewing equipment. Mass production begins by compression of variety to the point at which large numbers of artifacts form "equivalence classes" defined by common mechanical, electric and other tolerances. Failure to reduce variety, on the other hand, reduces

versatility, blocks the integration of systems and networks, and keeps them isolated and small in class extension. In planning "global" classification systems for knowledge, the not-so-hidden adversary is again the range in uncontrolled variety. The way to control is "standardization" -- of all kinds -- of symbols in the same language, between languages and most fundamentally (because hardest to do) of the conceptual classes with which the symbols are conditionally associated. Failure of "machine" translation is attributable to this last. The conceptual patterns formed from DSI's experienced at one latitude of the globe, and at one time in history, differ from those abstracted from different "geotemporal" locations. With systematic differences in abstract pattern, and accidental differences in syntactical structures, not only are exact word-for-word or sentence-for-sentence translations of natural languages impossible, but there must remain fundamental variety in conceptual pattern, equivalent to variety in meaning. This suggests a reorientation in such efforts, the prime objective being to determine the qualitative differences in concept patterns between language pairs, and the resulting limits on the "equivalence classes" that can or cannot be overcome. It is scarcely accidental that computers which could at first respond only to "machine" languages were incompatible. Computers gained compatibility in proportion to their capacity for "higher", more abstract languages, more flexible and versatile. Knowledge as a set of interrelated abstractions is most stable where most abstract. But any class at any level should be recognized for what it is -- an organic artifact -- evolving, adapting, growing. The stability of a class is "attacked" (disturbed) by new insights at all levels but especially at the lowest levels with unit extensions, SSI's. Greater intellectual and instrumental capability to discriminate is the motive power of science. But it produces increased variety in the form of interstitial interpolations of new classes. Thus any precoordinated system for describing knowledge crumbles, most rapidly at the bottom (close to the DSI's of new experiment) and less at the top. But no classes are permanently fixed.

In this light the stability of any particular subject area depends on several factors. One is the kinds of data (DSI's and SSI's) from which the class structures are abstracted. In the physical sciences, for example, there are very long, "repeatable" sets of data (SSI's). The physical sciences are those in which, by and large, variety is most easily suppressed, for they lack the enormous variety of living things. It is logical that man's ascendency over nature through the relations of science took place first in the physical sciences. Even here, concepts based on classes were highly unstable until certain higher levels were achieved, such as the insights of Copernicus, Keppler, Newton, and many others. Then with an intellectual "floor" of constraints on the variety of physical objects and force fields, the more complex structures of organisms and societies of organisms could be approached. Most resistant are those extremes of complexity which the cyberneticist calls the "very large system" -- large in terms of overwhelming variety, too large for more than broad, statistical homomorphisms. Conspicuous among sciences of large systems

are the social and behavioural sciences. Here the lengths of sequences of SSI's that constitute the basic data are still short, and the stability of classes correspondingly lower. Searchers for subjects in these areas experience this instability as a wider variety in nomenclature, different names for the same concept, or more confusing, the same names for slightly differing concepts. They look with envy on the conceptual stability of certain areas of the physical sciences. They should not be disturbed. They have to do with smaller class extensions, where, of course, the extension of a concept is always the finite number of SSI's that are naturally associated in the class. As the experiences of the field grow, the increasing extensions of these classes will push toward higher stability through the emergence of broader abstractions. These will apply more "universally", i.e., appear to be stable for more cases that are tested against their subjective thresholds of perception of differences in variety.

The same relative instability of class structure should occur in a field in which some kind of <u>uniqueness</u> is inherent in the subject. This is outstandingly the case for artifacts such as paintings, sculpture, or architectural features of buildings. These are unique, as are the critics' evaluations of them. There is a built-in uniqueness in a literary artistic or musical work in which, in spite of LD multiplication, the value of the work resides in the individual selection of the author, from the common store of variety such as the words of a language, the positions and colors and strokes of paint on a canvas, the spacing of musical notes.[19] Here again the secondary literature <u>about</u> the primary work, the criticism, abstract, or citation, has much irreducible variety which blocks suppression.

Faced, then, with improving access to the literature of the social and behavioural sciences and the arts and humanities, of which the common feature is some aspect of uniqueness, small classes, and wide variety -- what is the general direction that theory indicates our efforts should take? The answer should be clear, if we assume the ideas set forth above. We must consciously search out and suppress variety. Who will do this? The actual reduction of variety will be an immense task performed by thousands of workers. What devices will they use? Many are familiar, and have already been discussed: standardization of nomenclature, of indexing methods, of formats; establishment of intersystem compatibility; and special search strategies. One general result of practical experience should be stressed: to suppress variety has a cost. There is an inevitable "trade", a <u>quid pro quo</u> in every solution.

The theoretical pattern is simple. Because regulation involves variety and especially the flow of variety, and its flow anywhere on the IS path involves the same general elements, solving problems of access to knowledge involves the same general types of difficulty, same kinds of remedy, same general needs for exchange.

Let us close by pointing out that, as so often occurs, practice has preceded theory. Many or most of the devices mentioned above have been known, tried, used. Reasons for use resembled those suggested by the present hypothesis. An outstanding example is the class of device called thesaurus. Thesauri play essential roles in larger information search systems, especially in automated ones. While their greatest application has been to "man-machine interfaces", they are also used in "machine-machine interfaces" under other names such as "adapters" or "higher languages". The essential situation is that a large variety on one side of an interface cannot be tolerated on the other. In the case of large populations of users of an information system, the approach to the system is through "questions" which are structures of natural languages. All symbols, whether in "controlled" or "natural" languages, have been called here "conditionally" associated, (consistent with Pavlov's discovery that any random stimulus pattern can, if suitably reinforced, be firmly associated with a "naturally" associated pattern).[20] Naturally associated patterns range from automatic reflexes to higher concepts; in all of them systematic constraints are jointly imposed by body and environment. In natural language there are many sources of variety. One is the large number of synonyms and near-synonyms for the same or closely similar concepts (source of the original Roget). Another is the variety of codes for the same symbol (spellings), and in suffixes and prefixes to word stems. A third is the semantic dislocation by homonyms -- use of the same symbols for different concepts. All those and others can be present when access is by single or short compound terms. The variety is so great that an information system, particularly a machine in which access may be through a few codes or a single code, cannot respond without large variety -- e.g., low "recall" or low "precision", or both.[21]

A natural language thesaurus is used at interface c of the IS path to regulate the variety in segments cd of vocabulary among different searchers of an LD collection in their central segment bc. As explained, the variety regulated does not occur so much in the searchers' concepts (natural association) as in their access symbols (conditioned association). Let us simplify and assume a number of symbols, all equally likely to be employed by the population of searchers in access to a set of stored items. Suppose only one of these can actuate the access mechanism. Then the probability of retrieval of the set of items is 1/p without the thesaurus, close to 1 with it, if the p symbols nearly exhaust the variety in population access-vocabulary. Thesauri can be used with either pre- or post-coordination systems, but may have greater utility in the latter, since precoordination is equivalent to use of a syntactic ordering, which reduces variety. Thus a thesaurus would be particularly suitable in new fields, where new vocabulary is still being rapidly "coined", or that in use is having its "edges clipped" by free homonymity. The versatility of the searcher is increased, since achieving his goal is less dependent on his particular choice of index symbol. This stability is achieved by formation of a larger equivalence class. While this

much larger class extension has a slightly lower intension than any of its components, a single coordination, at most a few, rapidly restores the intension of the original concept. Thus there is enormous gain in system performance -- matching the man to the machine. A thesaurus is a "funnel" into which variety is poured hapharzardly, and out of which flows a thin, uniform stream which can be accurately directed into the narrow aperture of a small-neck "bottle" (machine or rigidly constrained system). It is an adapter. On a more abstract level, a higher language that endows computers and programs with the versatility of inter-communication and use is analogous to a thesaurus.

The cost, or exchange, is of several sorts. In postcoordination the searcher reserves greater freedom of "search strategy" but pays through performing more coordinations in order to achieve exhaustiveness (recall). The thesaurus cost includes that of the extra searches it makes possible. By far the greatest cost, however, is the task of setting up the thesaurus categories. These are best undertaken on a discipline-wide basis -- by all who expect to use the instrument. This suggests that in the social sciences and the humanities perhaps the most logical point of departure is to set up thesauri for the various subfields in profession-wide efforts. This would attack intra-subfield variety. At the same time, to assure eventual interfield compatibility, the various subfields should coordinate. The latter would involve careful comparison of concepts, and of terms used for access symbols. Obviously, the larger the area of knowledge, the greater the diversity and greater the work of suppressing variety in common concepts. Thousands of workers and many years may be required to achieve stability of access to knowledge in the social sciences and humanities. Fundamentally we are not removing variety from symbols so much as from ourselves -- in broadening the basis through shared knowledge. In this effort we may take satisfaction that cybernetics and information science are beginning to guide us. But we are not yet out of the woods. The light they shed is still feeble. It points out a direction, but not yet paths or their difficulty.

References and Notes

1. Ashby, W. Ross. An Introduction to Cybernetics. London, Chapman Hall, 1956, Chapter 4.

2. Heilprin, L.B. "Information Storage and Retrieval as a Switching System", Switching Theory in Space Technology, Aiken and Main, eds., Stanford University Press, 1963, 298-332.

3. Heilprin, L.B. and Goodman, F.L. "Analogy Between Information Retrieval and Education", American Documentation 16: July 1965, 163-169.

4. Heilprin, L.B. "Toward a Definition of Information Science", Automation and Scientific Communication, Proceedings 26th Annual Meeting, American Documentation Institute, Chicago, Ill., October 1963, 239-241.

5. Heilprin, L.B. "Critique and Response to Paper by Jesse H. Shera: An Epistemological Foundation for Library Science", The Foundations of Access to Knowledge, A Symposium, Montgomery, E.B., ed., Syracuse University Press, 1968, 26-35.

6. Heilprin, L.B. and Crutchfield, S.S. "Project Lawsearch: A Statistical Comparison of Coordinate and Conventional Legal Indexing", Parameters of Information Science, Proceedings 27th Annual Meeting, American Documentation Institute, Philadelphia, Pa., October, 1964, 215-234.

7. Heilprin, L.B. "On the Information Problem Ahead", American Documentation, 12, January, 1961, 6-14.

8. Ashby, W. Ross. Loc. cit., Chapter 11.

9. The LD propagation mode underlies communication of both cognitive and affective messages. A message can be regarded as a two-component vector (C,A). The messages of science have a larger C than A component, as measured by responses of the recipients; while those of the arts and humanities evoke a larger proportion of A to C response in the recipients. Neither component is ever completely absent.

10. Ashby, W. Ross. Loc. cit. Chapter 13, p. 247.

11. Bridgman, P.W. The Logic of Modern Physics, Macmillan, 1932, page 5.

12. Ashby, W. Ross. Loc. cit. Chapter 6

13. Random House Dictionary of the English Language, College edition, Random House, 1968, p. 1062.

14. Moliere, J.B.P. Le Bourgeois Gentilhomme, 1670. Act II, Sc 4.

15. Piaget, Jean Judgment and Reasoning in the Child. 1928 Translated, reprinted, Littlefield, Adams, 1966. Chapters IV,V.

16. Bruner, J.S., Goodnow, J.J., and Austin, G.A. A Study of Thinking. Wiley, 1956, p. 231.

17. Ashby, W. Ross Loc. cit. Chapter 2, p. 9.

18. Ashby, W. Ross Loc. cit. Chapter 7, pp. 130-4.

19. Heilprin, L.B. "Technology and the Future of the Copyright Principle", American Documentation 18, January 1968, pp. 6-11.

20. I.P. Pavlov Conditioned Reflexes, 1927 translated, Dover, 1960.

21. Cleverdon, C.W. "The Testing of Index Language Devices", ASLIB Proceedings, 15, Vol 4, April 1963.

22. Thanks are due Linda S. Handy for her long, patient care in typing this manuscript.